



The Statistical Cycle

Background

In this activity, you will use the stages of the statistical cycle to carry out an investigation and report on your findings. Your data must come from the 2012 CensusAtSchool.



STEP 1: The Problem

The first step when using the statistical cycle, is to decide on the topic of your investigation

e.g. what is the 'problem' you will investigate?

In this case your data must be numerical and compare the results of at least two distinct groups. For example:

Sex: Do boys have faster reaction times than girls?
Sexes (boys vs girls) are the distinct groups; reaction time is the numerical data



Breakfast: Do cereal eaters have better concentration than students who don't eat breakfast?

breakfast?

Breakfast types (cereal vs no breakfast) are the distinct groups; concentration time is the numerical data

Year level: Who has a longer foot length: your year level or another?
Year levels (yours vs another) are the distinct groups: foot length is the numerical data

- a)** Choose the problem you want to investigate. Now, write your investigation in the form of a question, like the examples above.

E.g. Do boys and girls grow the same rate from year 7 to year 10?

- b)** Define the key terms of your investigation question e.g. 'height' means height measured in cm without shoes. Explain what you already know about your chosen problem, and what you want to find out.

Height is recorded to the nearest cm and is without shoes. I know that as we get older we also get taller. This investigation will take a sample of students' heights for year 7 and year 10 and use summary statistics to see how much students grow in this time. Results can be used to see if the rate is the same for boys and girls.



STEP 2: The Plan

The second step in the cycle involves making the decisions about your data e.g. how big should your sample be? How will your sample be chosen? You also need to think about what you expect to find so that you are ready to talk about any unexpected results.

- c) What results do you expect to find? Write a sentence that predicts what you think the data will show you.

Answers will vary but should show that some thought has been put into this prediction e.g. Expect to see that boys grow at double the rate of girls.

- d) Look at the CensusAtSchool questionnaire and write down which questions provide the data you need to complete your investigation.

To answer this question a random sample containing data about sex and heights for my year level (Year 7) and for Year 10 is needed.

- e) What sample size will you use? Explain your choice.

About 20 students are needed in each group. A sample of 20 students is enough to be able to find any patterns in the data without it being too many to work with. The groups don't need to be exactly equal for comparisons to be made.

STEP 3: The Data

3

After defining your problem, and planning your investigation, the next step in the statistical investigation cycle is to organise the data so it is easy for you to interpret and for others to read.

- f) How did you collect and organise the data?

HINT: The data comes as an Excel spread sheet. You can use the 'Sort' function in the data tab to order the data e.g. by year level, sex, height. This will make it easier to display the data.

Ask for a random sample of 100 students. If there aren't about 20 students in each group, ask for another sample for that group. Data can be sorted by using the HINT above.

OR

Ask for 4 random samples of 20 for each group.

- g) Did you use all the data or did you "clean it" to eliminate any obvious errors? Explain why you did/did not clean the data. If you did clean the data, how did you decide what data to 'clean'?

Data can be cleaned to eliminate any obvious errors as outliers will affect the accuracy of results and how well any conclusions can be

applied to a larger population. In general all data should fall between 130 and 200 cm.

- h)** What level of accuracy is appropriate for this investigation? E.g. For very spread out data you may decide to arrange it so as to contain it in 6 - 10 groups. You may decide to round off your answers? Explain your decisions.

It is usual for the level of accuracy to be the same as for the original data i.e. to the nearest whole number

- i)** How did you display the data? Explain your choice.

Stem and leaf plots and dot plots are appropriate displays for Year 7. Stem and leaf plots allow the each data point to be retained and grouped. Dot plots are better used for grouped data or when the data is not spread over a large range such as here.

Year 7 Girls	
12	
13	1
14	7 8
15	0 0 3 4 4 5 5 5 6 7 7 9
16	0 0 1 2 5 5 6 8
17	

key 16|5 represents 165

Year 7 Boys	
12	
13	1
14	2 8 9
	0 1 1 3 4 5 5 5 5 6 6 8 8
15	9
16	0 0 1 1 2 3 5 5 6 9
17	0 1 8
18	6
19	

key 16|5 represents 165

Year 10 Girls	
13	
14	9
15	2 5 5 5 6 6 8 9
16	0 1 5 5 6 7 8 8 8 9
17	0 1 2 2
18	5
19	

key 16|5 represents 165

Year 10 Boys	
15	
16	5 8
17	3 3 3 4 5 5 6 9
18	0 0 0 0 0 5 5 9
19	6
20	

key 16|5 represents 165



STEP 4: The Analysis

Once you have sourced the data to answer your investigation 'problem', the next step in the statistical cycle is to analyse what the data tell you.

- j)** Produce and clearly display your data using summary statistics.

HINT: Summary statistics are best displayed in a table.

You will need to calculate the mean, median and range. You can do this in Excel by

using the formulae "=average" and "=median" to calculate the mean and median.

To calculate the mode use "=mode". You can choose to look at the mode or the modal range if your data are grouped.

nearest cm	Year 7				Year 10			
	mean	median	range	Modal group	mean	median	range	Modal group
Girls	156	156	37	150s	163	165	36	160s
Boys	159	158	55	150s	178	180	28	180s

nearest cm	Year 7		Year 10		Change	
	mean	median	mean	median	mean	median
Girls	156	156	163	165	7	9
Boys	159	158	178	180	19	22

Summary statistics should clearly show data related to the investigation



STEP 5: The Conclusion

This is the final step of the statistical investigation cycle. At this stage of the cycle, you are using your summary statistics to sum up your findings about your investigation. This is also where you communicate your findings using appropriate statistical vocabulary.

k) What is the answer to your original question? You will need to draw on your summary statistics from 'Step 4' to provide proof.

In answer to the original question it is clear that by both measures, namely mean and median, boys grow faster than girls from year 7 to year 10.

Mean height increase for girls is 7 cm while mean height increase for boys is 19 cm.

Median height increase for girls 9 cm for girls and 22 cm for boys.

A faster increase for boys is also shown by the change in modal group: from the 150cms to 160cms for girls and from the 150cms to the 180cms for boys.

- I) What meaning can be taken from your results? Can your answers be generalised beyond your sample? How could your results be applied in real life? Where should they not be used, or, used with caution?

The results show that, discounting growth spurts, in the 3 years from years 7 to year 10 girls grow about 2- 3 cm per year while boys grow about 6 - 7cm.

In this case the results can be applied to the population from which the sample was taken i.e. Australian students if no limiting parameters such as postcodes were set.

The theory of growth per year could be used to estimate the heights of students in year 8 and year 9. However the way that some students have growth spurts might make this prediction unreliable. Extrapolating the results to make predictions of heights of students below year 7 and above year 10 should be done with caution or not at all for the same reasons and because these groups of students might not grow at the same rates.